UNIVERSIDAD CARLOS III DE MADRID working papers

Working Paper
Economic Series 12 – 06
December 2012

Departamento de Economía
Universidad Carlos III of Madrid
Calle Madrid, 126
28903 Getafe, Madrid

"DIFFERENCES IN CITATION IMPACT ACROSS SCIENTIFIC FIELDS"

Juan A. Crespo[a], Yunrong Li[b], and Javier Ruiz-Castillo[b]

[a] Departamento de Economía Cuantitativa, Universidad Autónoma de Madrid
[b] Departamento de Economía, Universidad Carlos III

Abstract. This paper has two aims: (i) to introduce a novel method for measuring which part of overall citation inequality can be attributed to differences in citation practices across scientific fields, and (ii) to implement an empirical strategy for making meaningful comparisons between the number of citations received by articles in the 22 broad fields distinguished by Thomson Scientific. The paper is based on a model in which the number of citations received by any article is a function of the article's scientific influence, and the field to which it belongs. The model includes a key assumption according to which articles in the same quantile of any field citation distribution have the same *degree* of citation impact in their respective field. Using a dataset of 4.4 million articles published in 1998-2003 with a five-year citation window, we find that differences in citation practices between the 22 fields account for about 14% of overall citation inequality. Our empirical strategy for making comparisons of citation counts across fields is based on the strong similarities found in the behavior of citation distributions over a large quantile interval. We obtain three main results. Firstly, we provide a set of *exchange rates* to express citations in any field into citations in the all-fields case. (This can be done for articles in the interval between, approximately, the 71st and the 99th percentiles of their citation distributions). The answer is very satisfactory for 20 out of 22 fields. Secondly, when the raw citation data is normalized with our exchange rates, the effect of differences in citation practices is reduced to, approximately, 2% of overall citation inequality in the normalized citation distributions. Thirdly, we provide an empirical explanation of why the usual normalization procedure based on the fields' mean citation rates is found to be equally successful.

Acknowledgements. This is a third version of the paper published in this series in June 2012. The authors acknowledge financial support by Santander Universities Global Division of *Banco Santander*. Crespo and Ruiz-Castillo also acknowledge financial help from the Spanish MEC through grants SEJ2007-67436 and ECO2010-19596. Conversations with Jesús Carro, Eduardo Ley, Ignacio Ortuño, Juan Romo, Carlos Velasco, Ludo Waltman, and, above all, Pedro Albarrán, are deeply appreciated. All shortcomings are the authors' sole responsibility.

I. INTRODUCTION

The field dependence of reference and citation counts in scientific articles has been recognized since the beginning of Scientometrics as a field of study (see *inter alia* Pinski and Narin, 1976, Murugesan and Moravcsik, 1978, and Garfield, 1979). There are multiple reasons. Consider the differences across scientific disciplines in, for example, (i) size, measured by the number of publications in the periodical literature; (ii) the average number of authors per paper; (iii) the average paper length; (iv) the average number of papers per author in a given period of time; (v) the theoretical or experimental mix that characterizes each discipline; (vi) the average number of references per paper; (vii) the proportion of references that are made to other articles in the periodical literature; (viii) the percentage of internationally co-authored papers, or (ix) the speed at which the citation process evolves.

This paper develops a measuring framework where it is possible to quantify the importance of differences in citation practices. We use a model in which the number of citations received by an article is a function of two variables: the article's underlying scientific influence, and the field to which it belongs. In this context, the citation inequality of the distribution consisting of all articles in all fields –the all-fields case– is the result of two forces: differences in scientific influence, and differences in citation practices across fields. The first aim of the paper is how to isolate the citation inequality attributable to the latter, and how to measure its importance relative to overall citation inequality of all sorts.

The first difficulty we must confront is that the characteristics of the scientific influence distributions are *a priori* unknown. Thus, even if they were observable, we would not know how to compare the scientific influence of any two articles belonging to different fields. To overcome this difficulty, we make the strong assumption that articles in the same quantile of the scientific influence distribution have the same *degree* of scientific influence independently of the field to which they belong. Thus, if your article and mine belong, for example, to the 80th percentile of our respective distributions, then we assume that they have the same degree of scientific influence.

The next difficulty is that scientific influence is an unobservable variable. To overcome this difficulty, we may remain agnostic about the myriad of motives researchers have in their citation behavior as long as we are allowed to assume that citation impact varies monotonically with scientific influence (for a survey of the controversies concerning the meaning of citation counts, see Bornmann and Daniel, 2008). Thus, if one article has greater scientific influence than another one in the same homogeneous field, then we expect the former to have also a greater citation impact than the latter.[1] The monotonicity assumption ensures that, for any field, the quantiles of the (unobservable) scientific influence distribution coincide with the quantiles of the corresponding (observable) citation distribution. Therefore, if the mean citation of articles in the 80^{th} percentile of your field is, for example, twice as large as the mean citation of articles in the same percentile in my field, this means that your field uses double number of citations than mine to represent the same status in scientific influence. The implication is that the citation inequality observed at any quantile can be solely attributed to idiosyncratic differences in citation practices. Thus, the aggregation of this measure over all quantiles provides a method of quantifying the effect of these differences (This is, essentially, John Roemer's, 1998, model for the study of inequality of opportunities in an economic or sociological context).

We implement this model by using an additively decomposable inequality index, in which case the citation inequality attributed to differences in citation practices is captured by a between-group inequality term in the double partition by field and citation quantile (Ruiz-Castillo, 2003). Specifically, using a dataset of 4.4 million articles published in 1998-2003 with a five-year citation window and an appropriate citation inequality index, we estimate that the citation inequality attributable to differences in citation practices across the 22 fields distinguished by Thomson Scientific represents, approximately, 14% of overall citation inequality.

[1] The idea that citations is an observable indicator for a latent concept of scientific or scholarly influence, as well as the monotonicity assumption, are also found in Ravallion and Wagstaff (2011) in a different scenario: the construction of bibliometric measures of research impact.

It would appear that, regardless of how their impact can be measured, differences in publication and citation practices pose insurmountable obstacles to direct comparisons of the absolute number of citations received by articles in different fields. For example, in the dataset used in this paper, how can we interpret the fact that the mean citation in Mathematics is 2.4, about eight and a half times smaller than in Molecular Biology and Genetics where it is equal to 20.4 citations? This paper shows that the striking similarity between citation distributions (documented at different aggregation levels in Albarrán and Ruiz-Castillo, 2011, Albarrán *et al*., 2011, and Radicchi and Castellano, 2012a), causes the citation inequality attributable to different citation practices to be approximately constant over a wide range of quantiles. This allows its effect to be rather well estimated over that interval. Consequently, we provide a set of *exchange rates* and their standard deviations (StDevs hereafter) that serve to answer the following two questions. Firstly, how many citations in a given field are equivalent to, say, 10 citations in the all-fields case? For example, in *Clinical Medicine* the answer is 12.1 with a StDev of 0.6, while in *Mathematics* the answer is 3.3 with a StDev of 0.2. Secondly, how much can we reduce the effect of different citation practices by normalizing the raw citation data with the exchange rates? We find that this normalization procedure reduces this effect from 14% to, approximately, 2% of overall citation inequality.

The difficulty of comparing citation counts across scientific fields is a very well known issue that has worried practitioners of Scientometrics since its inception. Differences in citation practices are usually taken into account by choosing the world mean citation rates as normalization factors (see *inter alia* Moed *et al*., 1985, 1988, 1995, Braun *et al*., 1985, Schubert *et al*., 1983, 1987, 1988, Schubert and Braun, 1986, 1996, and Vinkler 1986, 2003). More recently, other contributions support this traditional procedure on different grounds (Radicchi *et al*., 2008, Radicchi and Castellano, 2012a, 2012b). In our last contribution, we find that using field mean citations as normalization factors leads practically to the same reduction of the effect of differences in citation practices on citation inequality as our exchange rates. We show how our model helps explaining why the traditional model is so successful.[2]

The rest of the paper consists of five Sections. Section II is devoted to a review of the literature on normalization using field citation means. Section III introduces the model for the measurement of the effect of differences in citation practices, while Section IV contains an estimate of this effect in term of an appropriate additively decomposable citation inequality index. Section V presents the estimation of average-based exchange rates and its StDevs over a large quantile interval, and discusses the consequences of using such field exchange rates and mean citations as normalization factors. Section VI contains some concluding comments.

II. A REVIEW OF THE LITERATURE

From an operational point of view, a scientific field is a collection of papers published in a set of closely related professional journals. A field is said to be *homogeneous* if the number of citations received by its papers is comparable independently of the journal where each has been published. The problem we confront in this paper arises when one wants to evaluate research units publishing in closely related but heterogeneous fields – such as a Chemistry department working in Organic and Inorganic Chemistry– or, more simply, when one wants to directly compare the citations received by two papers in different scientific fields at any aggregation level.

As indicated in the Introduction, the traditional solution is to rely on the world mean citation in each field as the normalization factor. Note that no confidence interval is usually provided in applications of this normalization procedure. This is probably due to the high variances that characterize highly skewed citation distributions (for the 22 fields covered in this paper, see column 4 in Table A in the Appendix). More importantly, no deep explanation is usually given for mean normalization. It is simply agreed that the field

[2] Methods that use mean citations or exchange rates as normalization factors belong to the class of target or "cited side" normalization procedures. Following an idea in Small and Sweeney (1985), source or "citing side" procedures have been recently suggested (see *inter alia* Zitt and Small, 2008, Moed, 2010, and Leydesdorff and Opthof, 2010). Since our dataset lacks citing side information, applying this type of procedure is beyond the scope of this paper.

mean citation captures well the expected value with which actual citation counts in that field can be related in order to compare normalized ratios across fields.

Let s_1 be the mean of a citation distribution, and let s_2 be the mean of those articles with citations above s_1. Under the idea that the difference $(s_2 - s_1)$ is a very good proxy for the StDev of citation distributions, Glänzel (2011) suggests a normalization of the raw data using this average-based difference.

In an important move, Radicchi *et al.* (2008) and Radicchi and Castellano (2012b) have recently justified the traditional solution on strong empirical grounds, namely, the *universality claim* according to which citation distributions in all fields exclusively differ by a scale factor. However, using a large dataset of 3.7 million articles published in 1998-2002, Albarrán *et al.* (2011) establish that the universality claim fails at both ends of the citation distributions at different aggregation levels, including a set of 219 sub-fields identified with the Web of Science subject-categories distinguished by Thomson Scientific (using a different methodology, Waltman *et al.*, 2011 reach the same conclusion). In the first place, Albarrán *et al.* (2011) find that the existence of a power law cannot be rejected at the top of the upper tail in 140 out of 219 sub-fields. On average, power laws represent 2% of all articles in a sub-field, and account for about 13.5% of all citations. However, the large dispersion of the power law parameters is a clear indication that excellence is not equally structured in all citation distributions.[3] In the second place, the proportion of articles without citations and with some citations below the mean at the sub-field level represent on average 24.7% and 43.9% of all articles, respectively, with large SDs equal to 13.9 and 12.5. Possibly, this is partly due to the fact that a common five-year citation window was taken for all sub-fields in spite of the large differences in the time that it takes for citation processes to reach a given degree of completion.

This assessment contrasts with the more optimistic view in Radicchi *et al.* (2008) that supports the universality claim with a methodology that does not inform about how to treat the assignment of articles to

[3] In addition, consider the possibility of defining a high-impact indicator over the sub-set of articles with citations above the 80th percentile of citation distributions. The distribution of high-impact values for the 219 sub-fields according to an indicator of this type is highly skewed to the right, and it presents some important extreme observations (see Herranz and Ruiz-Castillo, 2012).

multiple sub-fields, omits articles without citations, examines distributions at a limited set of points and, above all, covers only 14 of the 219 sub-fields. Radicchi and Castellano (2012b), which is free from other methodological shortcomings, focus only on 10 sub-fields within Physics. However, in a very important and more recent contribution that uses a dataset of about three million papers, covering 172 subject-categories, Radicchi and Castellano (2012a) –RC hereafter– also reject the universality claim. This seems to preclude certain normalization procedures. *"Making citation counts independent of the subject-categories seems therefore not possible with the use of linear transformations, because the difference between citation distributions of different subject-categories is not only due to a single scaling factor."* (RC, p. 2). More generally, *"A universal criterion for the complete suppression of differences among scientific domains probably does not exist. There are too many factors to account for, and consequently the 'philosophy' at the basis of a 'fair' normalization procedure is subjective"* (RC, p. 7). Nevertheless, RC demonstrate that, provided one is prepared to make strong assumptions, it is possible to find interesting normalization procedures. Ultimately, these normalization procedures work well in practice due to the similarity between citation distributions –a crucial aspect that deserves a few lines.

Generally, citation distributions are very different in many respects and, particularly, in size and mean citation rates. Consequently, it is very useful to use a size- and scale-invariant approach in order to focus on the shape of such distributions. One example is the Characteristic Scores and Scales (CSS hereafter) technique, introduced by Schubert *et al.* (1987) in the analysis of citation distributions. The CCS permits the partition of any citation distribution into a number of classes as a function of their members' citation characteristics. The following *characteristic scores* are determined: s_1 = mean of a citation distribution; s_2 = mean citation of articles with citations above s_1, and s_3 = mean citation of articles with citations above s_2. Although there is no universal distribution over the entire domain of all fields at any aggregate level, striking similarities over a broad partition of citation distributions at all aggregate levels have been found. In particular, on average, the proportion of articles at different aggregation levels that (i) receive none or few citations below s_1, (ii) are fairly well cited, namely, with citations between s_1 and s_2, and (iii) are remarkably or outstandingly cited with citations above s_2 is, approximately, 69/21/10. These three classes of articles account for the proportions 21/34/45 of all citations. The small StDevs that come with these average values establish the strong similarity between such highly skewed citation distributions (see Table 6 in Albarrán *et al.*, 2011a, and Figure 2 in Albarrán and Ruiz-Castillo, 2011).[4]

In this scenario, RC's scheme is based on the assumption that *"each discipline or field of research has the same importance for the development of scientific knowledge. A fair numerical indicator, based on citation numbers, must then assume values that do not depend on the particular scientific domain under consideration. Under this assumption, the probability to find a paper with a given value of the fair indicator must not depend on the discipline of the paper, or equivalently, the distribution of normalized citation counts must be the same for all disciplines."* (RC p. 7). RC's main result is that the transformation of raw citation numbers that makes the normalized citation distributions the same for all fields is a non-linear function that depends on only two parameters for every field: the mean, and an exponential factor that are rather stable over different publication years from 1980 to 2004. Moreover, mirroring the similarities between citation distributions just documented, RC find strong regularities: the exponential factor assumes approximately the same value for the vast majority of 172 subject-categories, suggesting that –after all– the main difference between the citation distributions of different subject-categories is given only by a scale factor. Consequently, the rescaling advocated in Radicchi *et al.* (2008) and Radicchi and Castellano (2012a) using simply the mean, despite not being strictly correct, seems a very good approximation of the transformation able to make citation counts not depending on the scientific domain.

III. THE MODEL

III. 1. Notation and Assumptions

Let N_f be the total number of articles in a homogeneous field f, and let $\mathbf{c}_f = (c_{f1},\ldots,c_{fNf})$ be the citation distribution for that field where, for each $i = 1,\ldots, N_f$, c_{fi} is the number of citations received by the i-th article.

[4] We saw before that articles receiving either no or few citations at the sub-field level had large StDevs. However, because of a strong negative correlation between these two groups, the broad class of poorly cited articles with citations below s_1 is located on average –as we have seen– around the 69th percentile of citation distributions with a StDev of 3.7.

Assume that there are F homogeneous fields, indexed by $f = 1,\ldots, F$. The total number of articles in the all-fields case is $N = \Sigma_f N_f$. The number of citations of any article, c_{fi}, is assumed to be a function of two variables: the field f to which the article belongs, and the scientific influence of the article in question, q_{fi}, which is assumed for simplicity to be a single-dimensional variable. Thus, for every f we write:

$$c_{fi} = \phi(f, q_{fi}),\ i = 1,\ldots, N_f. \tag{1}$$

Let $\boldsymbol{q}_f = (q_{f1}, q_{f2},\ldots, q_{fN_f})$ with $q_{f1} \leq q_{f2} \leq\ldots\leq q_{fN_f}$ be the ordered distribution of scientific influence in every field. It is important to emphasize that distribution $\boldsymbol{q}_f$ is assumed to be a characteristic of the field. Furthermore, no restriction is *a priori* imposed on distributions $\boldsymbol{q}_f$, $f = 1,\ldots, F$. Consequently, for any two articles i and j in two different fields f and g, the values q_{fi} and q_{gj} cannot be directly compared. To overcome this difficulty, in this paper we introduce some structure into the comparability problem by means of the following key assumption.

Assumption 1 (A1). *Articles at the same quantile* π *of any field scientific influence distribution have the same degree of scientific influence in their respective field.*

Typically, scientific influence is an unobservable variable. However, although the form of ϕ in Eq. 1 is unknown, we adopt the following assumption about it:

Assumption 2 (A2). *The function* ϕ *in expression (1) is assumed to be monotonic in scientific influence, that is, for every pair of articles i and j in field f, if* $q_{fi} \leq q_{fj}$, *then* $c_{fi} \leq c_{fj}$.

Under A2, the degree of scientific influence uniquely determines the location of an article in its field citation distribution. In other words, for every f, the partition of the scientific influence distribution $\boldsymbol{q}_f$ into $\boldsymbol{\Pi}$ quantiles of size $N_f/\boldsymbol{\Pi}$, $\boldsymbol{q}_f = (\boldsymbol{q}_f^{\boldsymbol{1}},\ldots, \boldsymbol{q}_f^{\boldsymbol{\pi}},\ldots, \boldsymbol{q}_f^{\boldsymbol{\Pi}})$, induces a corresponding partition of the citation distribution $\boldsymbol{c}_f = (\boldsymbol{c}_f^{\boldsymbol{1}}, \ldots, \boldsymbol{c}_f^{\boldsymbol{\pi}}, \ldots, \boldsymbol{c}_f^{\boldsymbol{\Pi}})$ into $\boldsymbol{\Pi}$ quantiles, where $\boldsymbol{c}_f^{\boldsymbol{\pi}}$ is the vector of the citations received by the $N_f/\boldsymbol{\Pi}$ articles in the π-th quantile of field *f*. Assume for a moment that we disregard the citation inequality within every vector $\boldsymbol{c}_f^{\boldsymbol{\pi}}$ by assigning to every article in that vector the mean citation of the vector itself, namely, $\mu_f^{\boldsymbol{\pi}}$. Since the quantiles of citation impact correspond –as we have already seen– to quantiles of the underlying scientific influence distribution, holding constant the degree of scientific influence at any level as in A1 is equivalent to holding constant the degree of citation impact at that level. Thus, the interpretation of the fact that, for example, $\mu_f^{\boldsymbol{\pi}} = 2\ \mu_g^{\boldsymbol{\pi}}$ is that, on average, field *f* uses twice the number of citations as field *g* to represent the same underlying phenomenon, namely, the same degree of scientific influence in both fields. Hence, for any π, the difference between $\mu_f^{\boldsymbol{\pi}}$ and $\mu_g^{\boldsymbol{\pi}}$ for articles with the same degree of scientific influence is entirely attributable to differences in citation practices between the two fields.

Welfare economists would surely recognize the above as Roemer's (1998) model for the inequality of opportunities where individual incomes (or other indicators of performance, such as educational outcomes) are assumed to be a function of two types of factors: a set of variables outside an individual's responsibility –the *circumstances*, mainly inherited from our parents–, and *effort*, an unobservable single dimensional variable entirely within the sphere of each individual's responsibility. Circumstances allow a partition of the population into *types*. The distribution of effort within each type is assumed to be a characteristic of the type. Consequently, the amounts of effort exercised by individuals from different types are not ethically comparable. However, degrees of effort, measured by quantiles of the effort distribution for each type, are assumed to be comparable (A2). Under the monotonicity assumption (A1), quantiles of effort are seen to correspond with observable income quantiles. In this model, income inequality holding constant the degree of effort by every type is seen to be entirely due to differences in circumstances, or to the *inequality of* *opportunities* at this degree of effort. Income inequality due to differences in effort is not worrisome from a social point of view. It is income inequality due to differences in circumstances, namely, the inequality of opportunities, what society might attempt to compensate for. Individuals are articles, the equivalent of income is citations, types are fields, and effort is scientific influence.

III.2. The Measurement of the Effect of Differences in Citation Practices

For any population partition, we are interested in expressing the overall citation inequality as the sum of two terms: a weighted sum of *within-group* inequalities, plus a *between-group* inequality component. An inequality index is said to be *decomposable by population subgroup*, if the decomposition procedure of overall inequality into a within-group and a between-group term is valid for any arbitrary population partition. In the relative, or scale-invariant inequality case it is customary to calculate the between-group component by applying the inequality index to a citation vector in which each article in a given subgroup is assigned the subgroup's citation mean. Under this convention, it is well known that the Generalized Entropy (GE hereafter) family of inequality indices are the only measures of relative inequality that satisfy the usual properties required from any inequality index[5] and, in addition, are decomposable by population subgroup (Bourguignon, 1978, and Shorrocks, 1980, 1984). Without loss of generality, it is useful to develop the following measurement framework in terms of only one member of this family, the first Theil index, denoted by I_1. For any citation distribution $\boldsymbol{Q} = (c_1, \ldots, c_l, \ldots, c_N)$ with N articles indexed by $l = 1, \ldots, N$, the citation inequality index I_1 is defined as:

$$I_1(\boldsymbol{Q}) = (1/N)\, \Sigma_l\, (c_l/\mu) \log (c_l/\mu), \tag{2}$$

where μ is the mean of distribution $\boldsymbol{Q}$.

[5] Namely, continuity; scale invariance; invariance to population replications, or size-invariance, and S-convexity that ensures that transfers from an article with more citations to another with fewer citations without altering their ranking reduces citation inequality.

Let $\boldsymbol{C} = \cup_f \boldsymbol{c}_f$ be the overall citation distribution in the all-fields case. For each π, define the vector $\boldsymbol{c}^{\boldsymbol{\pi}} = (\boldsymbol{c}_1^{\boldsymbol{\pi}}, \ldots, \boldsymbol{c}_f^{\boldsymbol{\pi}}, \ldots, \boldsymbol{c}_F^{\boldsymbol{\pi}})$ of size $(\Sigma_f N_f)/\Pi = N/\Pi$. Clearly, $\boldsymbol{C} = (\boldsymbol{c}^{\boldsymbol{1}}, \ldots, \boldsymbol{c}^{\boldsymbol{\pi}}, \ldots, \boldsymbol{c}^{\boldsymbol{\Pi}})$, and the set of vectors $\boldsymbol{c}^{\boldsymbol{\pi}}$, $\pi = 1, \ldots, \Pi$, form a partition of $\boldsymbol{C}$. The formula for the I_1 index when written in decomposable form for the partition $\boldsymbol{C} = (\boldsymbol{c}^{\boldsymbol{1}}, \ldots, \boldsymbol{c}^{\boldsymbol{\pi}}, \ldots, \boldsymbol{c}^{\boldsymbol{\Pi}})$ is the following

$$I_1(\boldsymbol{C}) = \Sigma_\pi v^\pi I_1(\boldsymbol{c}^{\boldsymbol{\pi}}) + I_1(\boldsymbol{\mu}^{\boldsymbol{1}}, \ldots, \boldsymbol{\mu}^{\boldsymbol{\Pi}}), \tag{3}$$

where v^π is the share of total citations received by articles in vector $\boldsymbol{c}^{\boldsymbol{\pi}}$, and $I_1(\boldsymbol{\mu}^{\boldsymbol{1}}, \ldots, \boldsymbol{\mu}^{\boldsymbol{\Pi}})$ is the citation inequality of the distribution $\boldsymbol{m} = (\boldsymbol{\mu}^{\boldsymbol{1}}, \ldots, \boldsymbol{\mu}^{\boldsymbol{\Pi}})$ in which each article in a given vector $\boldsymbol{c}^{\boldsymbol{\pi}}$ is assigned the vector's citation mean, $\mu^\pi = \Sigma_f [(N_f/N]\mu_f^\pi$. Next, for each π, the decomposability property of I_1 is applied to the partition into F fields, $\boldsymbol{c}^{\boldsymbol{\pi}} = (\boldsymbol{c}_1^{\boldsymbol{\pi}}, \ldots, \boldsymbol{c}_F^{\boldsymbol{\pi}})$:

$$I_1(\boldsymbol{c}^{\boldsymbol{\pi}}) = \Sigma_f v_f^\pi I_1(\boldsymbol{c}_f^{\boldsymbol{\pi}}) + I_1(\boldsymbol{\mu}_1^{\boldsymbol{\pi}}, \ldots, \boldsymbol{\mu}_F^{\boldsymbol{\pi}}), \tag{4}$$

where v_f^π is the share of citations in vector $\boldsymbol{c}^{\boldsymbol{\pi}}$ received by articles in quantile $\boldsymbol{c}_f^{\boldsymbol{\pi}}$, and $I_1(\boldsymbol{\mu}_1^{\boldsymbol{\pi}}, \ldots, \boldsymbol{\mu}_F^{\boldsymbol{\pi}})$ is the citation inequality of the distribution in which each article in quantile π of field f receives that sub-group's mean citation $\boldsymbol{\mu}_f^{\boldsymbol{\pi}}$. Inserting (4) into (3), overall citation inequality is seen to be:

$$I_1(\boldsymbol{C}) = W + S + IDCP, \tag{5}$$

where:

$$W = \Sigma_\pi v^\pi \Sigma_f v_f^\pi I_1(\boldsymbol{c}_f^{\boldsymbol{\pi}}) = \Sigma_\pi \Sigma_f v^{\pi,f} I_1(\boldsymbol{c}_f^{\boldsymbol{\pi}})$$

$$S = I_1(\boldsymbol{\mu}^{\boldsymbol{1}}, \ldots, \boldsymbol{\mu}^{\boldsymbol{\Pi}})$$

$$IDCP = \Sigma_\pi v^\pi I_1(\boldsymbol{\mu}_1^{\boldsymbol{\pi}}, \ldots, \boldsymbol{\mu}_F^{\boldsymbol{\pi}}),$$

where $v^{\pi,f}$ is the share of total citations received by articles in vector $\boldsymbol{c}_f^{\boldsymbol{\pi}}$. The term W in equation (5) is a within-group term, which captures the weighted citation inequality within each quantile in every field. Obviously, since all articles in each vector $\boldsymbol{c}_f^{\boldsymbol{\pi}}$ belong to the same field, there is no difficulty in computing the expression $I_1(\boldsymbol{c}_f^{\boldsymbol{\pi}})$. Clearly, for large $\boldsymbol{\Pi}$, $I_1(\boldsymbol{c}_f^{\boldsymbol{\pi}})$, and hence term W is expected to be small. The term S is the citation inequality of the distribution $\boldsymbol{m} = (\boldsymbol{\mu}^{\boldsymbol{1}},..., \boldsymbol{\mu}^{\boldsymbol{\Pi}})$ in which each article in a given quantile $\boldsymbol{\pi}$ is assigned the quantile's citation mean, $\mu^{\pi} = \Sigma_f\ [(N_f/N]\mu_f^{\pi}$. Thus, S is a measure of citation inequality at different degrees of citation impact that captures well the skewness of science in the all-fields case. Due to the high skewness of all citation distributions (see *inter alia* Albarrán and Ruiz-Castillo, 2011, and Albarrán *et al.*, 2011), the term S is expected to be large. Finally, for any π, the expression $I_1(\boldsymbol{\mu}_1^{\boldsymbol{\pi}},..., \boldsymbol{\mu}_F^{\boldsymbol{\pi}})$, abbreviated as $I(\boldsymbol{\pi})$, is the citation inequality attributable to differences in citation practices according to I_1. Thus, the weighted average that constitutes the third term in Eq. 5, denoted by *IDCP* (*I*nequality due to *D*ifferences in *C*itation *P*ractices), provides a good measure of the citation inequality due to such differences.

IV. THE ESTIMATION OF THE EFFECT OF DIFFERENCES IN CITATION PRACTICES

IV. 1. The Data

Since we wish to address a homogeneous population, in this paper only research articles or, simply, articles, are studied. The dataset consists of 4.4 million articles published in 1998-2003, and the 35 million citations they receive after a common five-year citation window for every year, namely, citations received from 1998 to 2002 for articles published in 1998, up to 2003 to 2007 for articles published in 2003.

Since in this paper we must work with partitions of the N articles, we identify the set of homogeneous fields with the 20 broad fields for the natural sciences and two for the social sciences distinguished by Thomson Scientific. Table A in the Appendix presents the number of articles and mean citation rates. For convenience, fields are classified in terms of four large groups: Life Sciences, Physical Sciences, Other Natural Sciences, and Social Sciences, which represent, respectively, 40.4%, 28.7%, 25.7%, and 5.2% of all articles.

IV. 2. The Choice of Inequality Index and the Number of Quantiles

The GE family can be described by means of the following convenient cardinalization:

$$I_a(\mathbf{C}) = (1/N)\,(1/a^2 - a)\,\Sigma_l\,(c_l/\mu^a - 1),\ a \neq 0,1; \tag{6}$$

$$I_0(\mathbf{C}) = (1/N)\,\Sigma_l \log(\mu/c_l);$$

$$I_1(\mathbf{C}) = (1/N)\,\Sigma_l\,(c_l/\mu)\log(c_l/\mu).$$

Parameter a summarizes the sensitivity of I_a in different parts of the productivity distribution: the more positive (negative) a is, the more sensitive I_a is to differences at the top (bottom) of the distribution (Cowell and Kuga, 1981). I_1 is the original Theil index, while I_0 is the mean logarithmic deviation. Consider any partition of $\mathbf{C}$ into, say, K subgroups, indexed by $k = 1,\ldots, K$, $\mathbf{C} = (\mathbf{c}^1,\ldots, \mathbf{c}^K)$. The formula for the GE index when written in decomposable form is the following:

$$I_a(\mathbf{C}) = \Sigma_k\, w_a^{\,k}\, I_a(\mathbf{c}^k) + I_a(\boldsymbol{\mu}^1,\ldots, \boldsymbol{\mu}^K), \tag{7}$$

where $w_a^{\,k} = [(v^k)^a\,(p^k)^{1-a}]$; v^k is the share of total citations held by articles in subgroup k; p^k is sub-group k's population share, and $I_a(\boldsymbol{\mu}^1,\ldots, \boldsymbol{\mu}^K)$ is the between-group inequality calculated as if each article in subgroup k received that sub-group's mean citation $\boldsymbol{\mu}^k$. In particular, for the partition of distribution $\mathbf{C}$ into $\boldsymbol{\Pi}$ quantiles, $\mathbf{C} = (c^1,\ldots, c^{\boldsymbol{\pi}}, \ldots, c^{\boldsymbol{\Pi}})$, we have:

$$I_a(\mathbf{C}) = \Sigma_\pi\, w_a^{\,\pi}\, I_a(\mathbf{c}^{\boldsymbol{\pi}}) + I_a(\boldsymbol{\mu}^1,\ldots, \boldsymbol{\mu}^{\boldsymbol{\Pi}}). \tag{8}$$

In order to select some member of the GE family of inequality indicators, we may take into account the following three considerations. Firstly, the weights in the within-group term in expression (7), $w_a^{\,k}$, add up to one only for $a = 0$ and $a = 1$. In any other case, the within-group term will not be a weighted average of the sub-group values $I_a(\mathbf{c}^k)$. More importantly, it can be shown that $1 - \Sigma_k w_a^k$ is proportional to the between-group term in (7). This leads to serious difficulties of interpretation of the decomposition in question (see Shorrocks, 1980). Secondly, the behavior of the members of the family when $a \geq 2$ are rather extreme: they show increasingly little concern for transfers except among the very highly cited articles (see also Shorrocks, 1980). In highly skewed distributions this can be problematic. For example, Albarrán *et al.* (2012) show that the elimination of the most highly cited article in each of the 22 fields, that is to say, 22 articles among a dataset of 4.4 million, reduces citation inequality by more than 5%. Thirdly, in the case $a = 2$, for example, the weights become $w_2^k = (\mu^k/\mu)\ v^k$. In particular, for the partition into Π quantiles in expression (8), we have $w_2^\pi = (\mu^\pi/\mu)\ v^\pi$. Thus, for high values of π, w_2^π becomes very high indeed. As we will presently see, the last two facts imply that most of the *IDCP* term is accounted for by the last few quantiles.

These considerations advise choosing either $a = 0$ or $a = 1$. In the first case, for every π $w_0^\pi = p^\pi = N/\Pi$, that is, w_0^π is quantile's π demographic share. Instead, $w_1^\pi = v^\pi$, the share of citations in quantile π relative to total citations. In economics, the demographic weighting by p^π when $a = 0$ is usually preferred on normative grounds. In our context, the choice $a = 1$ seems more appropriate, in which case the higher the quantile π, the greater the weight v^π assigned to $I_1(\mu_1^\pi, \ldots, \mu_F^\pi)$ in the *IDCP* term in Eq. 5. The problem with this choice (as well as in the case $a = 0$), is that there is a considerable percentage of articles in all fields that receive zero citations, and the index I_1 (as well as I_0) in (6) is only defined for positive numbers. Therefore, we experimented with the following options: assigning to articles without citations the values $\varepsilon_1 = 0.1$, and $\varepsilon_2 = 0.01$ whenever $a = 0, 1$, or adopting the convention $0 \log(0) = 0$ for these articles in the case $a = 1$. Since we must decide at the same time on the value of Π, we have estimated Eq. 5 for the following choices: (i) $a = 2$;

(ii) $a = 0$ and $\varepsilon_1 = 0.1$; (iii) $a = 0$ and $\varepsilon_2 = 0.01$; (iv) $a = 1$ and $\varepsilon_1 = 0.1$; (v) $a = 1$ and $\varepsilon_2 = 0.01$, and (vi) $a = 0$ and $0 \log(0) = 0$, on the one hand, and $\Pi = 10$, 50, 100, and 1,000 on the other hand. The results are in Table 1.

Table 1 around here

The following three points should be noted. Firstly, when $a = 2$ the quantile choice affects the relative importance of the three terms in decomposition (5). As Π grows, the vectors c_f^{π} for all f become smaller and smaller. Consequently, the within-group citation inequality term W loses importance in favor of the S term. Unfortunately, the *IDCP* term is also pretty sensitive to the quantile choice. Secondly, when $a = 0$ a similar pattern is observed, with two differences: the term W is very small indeed, and the sensitivity of the *IDCP* term to Π is smaller than when $a = 2$. However, being very sensitive to transfers at the lower tail of citation distributions, the importance of the *IDCP* term according to I_0 is generally higher than for the other two choices of parameter a, and dramatically increases when we go from ε_1 to ε_2. Thirdly, when $a = 1$ the *IDCP* term remains essentially constant for all choices of Π and ε, ranging from 13.22% to 13.95%. Moreover, the order of magnitude of the *IDCP* term is similar to the case $a = 2$. Thus, we decide to stick to the following choices: $a = 1$, $0 \log(0) = 0$, and $\Pi = 1{,}000$.

V. COMPARABILITY AND NORMALIZATION RESULTS

This Section analyzes two empirical problems: (i) how to compare the citations received by two articles in any pair of the 22 fields in our dataset by using exchange rates that are approximately constant over a large quantile interval, and (ii) how much the effect of differences in citation practices is reduced when these exchange rates, or the field mean citations are used as normalization factors.

V. 1. The Comparison of Citation Counts Across Different Fields

Mean citations of comparable articles belonging to the same quantile can be used to express the citations in any field in terms of the citations in a reference situation. For example, if we let μ^{π} be the mean citation of all articles in quantile π, then the *exchange rates at quantile* π, $e_f(\pi)$, defined by

$$e_f(\pi) = \mu_f^{\pi}/\mu^{\pi}, \tag{9}$$

can be seen to answer the following question: how many citations for an article at the degree π of scientific influence in field f are equivalent on average to one citation in the all-fields case? In the metaphor according to which a field's citation distribution is like an income distribution in a certain currency, the exchange rates $e_f(\pi)$ permit to express all citations in the same reference currency for that π: since c_{fi} is the number of citations received by article i in quantile π of field f, the ratio $c_{fi}^*(\pi) = c_{fi}/e_f(\pi)$ is the equivalent number of citations in the reference currency at that quantile. Naturally, if for many fields $e_f(\pi)$ were to drastically vary with π, then we might not be able to claim that differences in citation practices have a common element that can be precisely estimated. However, we next establish that exchange rates are sufficiently constant over a wide range of quantiles.

It is very instructive to have a graphical representation of how the effect of differences in citation practices, measured by $I(\pi)$, changes with π when $\Pi = 1{,}000$ (since $I(\pi)$ is very high for $\pi < 600$, for clarity these quantiles are omitted from Figure 1). It is observed that $I(\pi)$ is particularly high until $\pi \approx 700$, as well as for a few quantiles at the very upper tail of citation distributions. However, $I(\pi)$ is strikingly similar for a wide range of intermediate values.[6] In this situation, it is reasonable to define an average-based *exchange rate* (*ER* hereafter) over some interval $[\pi_m, \pi^M]$ in that range as

$$e_f = [1/(\pi^M - \pi_m)] [\Sigma_\pi e_f(\pi)]. \tag{10}$$

An advantage of this definition is that we can easily compute the associated StDev, denoted by σ_f. The fact that, for each *f*, the $e_f(\pi)$ defined in (9) are very similar for all π in the interval $[\pi_m, \pi^M]$ would manifest itself in a small σ_f, and hence in a small coefficient of variation $CV_f = \sigma_f / e_f$.

Figure 1 around here

We find that the choice $[\pi_m, \pi^M] = [706, 998]$ –where $I(\pi)$ for most π is equal to or smaller than $I(\pi_m)$ = 0.1081 and $I(\pi^M)$ = 0.1084– is a good one. The *ER*s e_f, as well as the σ_f, and CV_f are in columns 1 to 3 in Table 2. For convenience, *ER*s are multiplied by 10. Thus, for example, the first row indicates that 15.8 citations with a StDev of 0.9 for an article in Biology and Biochemistry between, approximately, the 71st and the 99th percentile of its citation distribution, are equivalent to 10 citations for an article in that interval in the all-sciences case. We find it useful to divide fields into three groups according to the CV_f. Group I (colored in green in Table 1), consisting of 10 fields, has a CV_f smaller than or equal to 0.05. This means that the StDev of the exchange rate is less than or equal to five percent of the exchange rate itself. Hence, we consider *ER*s in this group as highly reliable. Group II (black), consisting of 10 fields, has a CV_f between 0.05 and 0.10. We consider *ER*s in this group as fairly reliable. Group III (red), consists of two fields: *Computer Science*, with a CV_f greater than 0.10, which is known from previous work to behave as an outlier (Herranz and Ruiz-

[6] It is important to emphasize that this is consistent with the stylized facts characterizing citation distributions discussed in Section II and documented in Albarrán and Ruiz-Castillo (2011), and Albarrán *et al.* (2012): although the percentages of articles belonging to three broad classes are very similar across fields, citation distributions are rather different in a long lower tail and at the very top of the upper tail.

Castillo, 2012), and the *Multidisciplinary* field with a CV_f greater than 0.15, a hybrid field that does not behave well either in RC. The results for these two fields should be considered unreliable.

Table 2 around here

As is observed in column 4 in Table 2, on average the interval [706, 998] includes 72.1% of all citations (with a StDev of 3.9). Expanding the interval in either direction would bring a larger percentage of citations. It turns out that the *ER*s do not change much. However, they exhibit greater variability. For example, moving the upper bound π^M to quantile 1,000 would increase the percentage of citations to 76.3% (StDev = 5). However, the CV_f would increase in all but three fields, and the number of fields in Group I would decrease from 10 in the reference case down to 8. In the other direction, moving the lower bound π_m to quantiles 700, or 694, for example, would slightly increase the percentage of citations to 72.7%, (StDev = 3.8) and 73.3% (StDev = 3.8). However, relative to the initial choice, in these two instances the CV_f would increase in 13 out of 22 fields, and the number of fields in Groups I would decrease from 10 to 9. On the other hand, after normalization by the *ER*s corresponding to the three alternatives [706, 1000], [700, 998], and [694, 998], the *IDCP* term represents essentially the same percentage of the overall citation inequality in the normalized distributions (see below). Therefore, we retain the interval [706, 998] in the sequel.

V. 2. Normalization Results

Overall citation inequality due to differences in scientific influence –captured by the *W* and *S* terms in Eq. 6– is not worrisome. Instead, we would like to eliminate as much as possible the citation inequality attributable to differences in citation practices. Thus, the impact of any normalization procedure can be evaluated by the reduction in the term *IDCP* before and after normalization. Figure 2 focuses on the product $v^{\pi} I(\pi)$ as a function of π. Of course, the term *IDCP* is equal to the integral of this expression (for clarity, quantiles $\pi < 600$, and $\pi > 994$, are omitted from Figure 2). Note the strong effect of the weights v^{π} as π

increases. As a matter of fact, the percentage of *IDCP* reached at π = 400, 700, 900, and 990 are 15.2%, 35.9%, 61.9%, and 88.9%, respectively.[7]

Relative to the blue curve, the red curve illustrates the correction achieved by normalization: the size of the *IDCP* term is very much reduced. The numerical results before and after this normalization are in Panels A and B in Table 3. Note that, as before, the term *W* is small, while the term *S* is large. Both terms remain essentially constant after normalization. However, in absolute terms the *IDPC* term is reduced from 0.1221 to 0.0167, a 86.3% difference. Of course, total citation inequality after normalization is also reduced. On balance, the *IDPC* term after normalization only represents 2.09% of total citation inequality –a dramatic reduction from the 13.95% with the raw data.

Table 3 and Figure 2 around here

However, it should be recognized that in the last two quantiles and, above all, in the [1, 705] interval normalization results quickly deteriorate.[8] It would appear that a convenient alternative consists of normalizing the lower tail of the original distributions by some appropriate *ER*s within the [1, 705] interval. The problem is that citation inequality due to different citation practices in that interval is both high and extremely variable for different quantiles. It turns out that the *ER*s computed according to equation (10) for the entire [1, 705] interval lead to a worsening of the situation. However, when we restrict ourselves to the interval [356, 705] we are able to improve matters somewhat. The new *ER*s, together with their high σ_f, and CV_f, are in Table B in the Appendix. The second set of *ER*s is rather different: only in seven cases do they stay within one StDev of the first set in Table 2. On the other hand, CV_fs increase so much that seven fields are now in Group IV when we only had one field in that group before. Be that as it may, the end result is that

---

7 Being very sensitive to transfers at the upper tail of citation distributions, the percentage of *IDCP* reached at π = 400, 700, 900, and 990 according to I_2 are very much lower than according to I_1: 0.7%, 4.7%, 17.1%, and 50%, respectively. Thus, half of the *IDCP* is accounted for by the last ten quantiles.

8 It should be noted that RC also find that the non-linear transformation that makes the normalized citation distributions the same for all fields (see Section II) becomes less descriptive beyond the top 10% of highly cited articles, and the removal of the bias in the raw data worsens.

after normalization by the two sets of *ER*s the *IPC* only goes down to 1.86% of total citation inequality (see Panel C in Table 3) versus 2.09% with a single set of *ER*s. Most of this figure, or 1.36%, is still accounted for by what happens in the interval [1, 705]. We must conclude that the improvement over the alternative with a single set of *ER*s is, at most, very slight.

As indicated in the Introduction and discussed in Section II, the difficulties of combining heterogeneous citation distributions into broader aggregates have been traditionally confronted using mean citations as normalization factors. In our dataset, the *IDCP* term after the traditional normalization procedure only represents 2.05% of total citation inequality (see Panel D in Table 3). The two solutions are so near that we refrain to illustrate the latter in Figure 2 because it will be indistinguishable with the red curve after normalization by our *ER*s. This confirms the results in RC, where it is concluded that, despite not being strictly correct, this procedure is a very good approximation of the two-parameter transformation able to make citation counts independent of the scientific field.

The question is, how can this similarity of results be accounted for? The explanation is as follows. As documented in Albarrán *et al.* (2011), field mean citations μ_f are reached, on average, at the 69.7 percentile with a StDev of 2.6, that is, at the lower bound of the [706, 998] interval. Thus, the *ER*s based on mean citations, $e_f(\mu_f) = \mu_f/\mu$ (reproduced in column 5 in Table 1), are approximately equal our own *ER*s (in column 1 in that Table). In other words, let μ'_f and μ' be the mean citations in each field and the population as a whole restricted to the [706, 998] interval, and consider the average-based *ER*s based on these restricted means: $e_f(\mu'_f) = \mu'_f/\mu'$ (see column 6 in Table 1). Since field citation distributions differ approximately by a set of scale factors only in the interval [706, 998], these scale factors should be well captured by any average-based measure of what takes place in that interval –such as our own e_f, or the new $e_f(\mu'_f)$. However, the latter *ER*s are essentially equal to the old ones, that is, for each f, $e_f(\mu'_f) \approx e_f(\mu_f) \approx e_f$.

Finally, we have estimated the reduction in the *IDCP* term when, following Glänzel (2011), the normalization factors are made equal to the difference ($s_2 - s_1$) for each field, where s_1 and s_2 are the first two scores in the Characteristic Scores and Scales approach discussed in Section II. The results are in Panel E of Table 3. Interestingly, for the last two quantiles the reduction is larger than in all previous cases. However, the entire *IDCP* term after this third normalization becomes 3% –rather than 2%– of overall citation inequality.

VI. CONCLUSIONS

The lessons that can be drawn from this paper can be summarized in the following four points.

1. We have provided a simple method for the measurement of the effect of differences in citation practices across scientific fields. Using a member of a family of additively separable citation inequality indices, this effect is well captured by a between-group term –denoted *IDCP*– in the double partition by field and quantile of the overall citation distribution in the all-fields case. It should be noted that this is a distribution free method, in the sense that it does not require that the scientific influence or the citation distributions satisfy any specific assumptions. Using a large dataset of 4.4 million articles in 22 scientific fields and a five-year citation window, we have estimated that the *IDCP* term represents about 14% of overall citation inequality –a result independent of the number of quantiles.

2. The striking similarity of citation distributions allows the effect of idiosyncratic citation practices to be rather well estimated over a wide range of intermediate quantiles where citation distributions seem to differ by a scale factor. Consequently, a set of *ER*s has been estimated in the interval $[706, 998]$ for two purposes: the translation of citation counts of articles in different fields within that interval into the citations in a reference situation, and the normalization of the raw citation data. Such *ER*s are estimated with a reasonably low StDev for 20 out of 22 fields.

It should be stressed that, for uncited and poorly cited articles below the mean, and for articles in the very upper tail of citation distributions, no clear answer to the comparability of citation counts for articles in different fields can be provided. Since the citation process evolves at different velocity in different fields, using variable citation windows to ensure that the process has reached a similar stage in all fields should improve field comparability at the lower tail of citation distributions. Naturally, we may also worry about how to compare citation counts in the last two quantiles of citation distributions. Given the fact that in this key segment the citation impact appears to be very diverse across fields, perhaps this task should not be even attempted. Until we know more concerning how differential citation practices operate in these top quantiles, the most we can do within this paper's framework is to use *ER*s $e_f(\pi)$ for $\pi = 999, 1000$.

3. The success of any normalization procedure in eliminating as much as possible the impact of differences in citation practices can be evaluated by the reduction it induces in the *IDCP* term. In our case, it has been established that both the procedure that uses our *ER*s, as well as the traditional method of taking the field citation means as normalization factors reduces the importance of the *IDCP* term relative to overall citation inequality from, approximately, 14% to 2%. The paper provides an empirical explanation of why the two methods are equally successful. Finally, we estimate that the normalization advocated in Glänzel (2011) reduces the *IDCP* term to 3% of overall citation inequality.

Other normalization proposals –such as the one in RC, or those based on citing side procedures quoted in the Introduction, might be analogously evaluated. In turn, it would be interesting to evaluate the normalization procedure based on the *ER*s in terms of the reduction of the bias in the RC model. Given how near our *ER*s are from those based on the fields' mean citation rates, the conjecture is that our procedure would perform as well as the approximation provided by these means in RC.

4. Policy makers and other interested parties should be very cautious when comparing citation performance in different scientific fields. More research is still needed. In particular, we need to study the robustness of our strategy to other datasets, as well as to extend it to lower aggregation levels. However, together with the important contribution by RC, the results of this paper indicate that the combination of interesting assumptions with the empirical similarity of citation distributions paves the way for meaningful comparisons of citation counts across heterogeneous fields.

REFERENCES

Albarrán, P. and J. Ruiz-Castillo (2011), "References Made and Citations Received By Scientific Articles", *Journal of the American Society for Information Science and Technology*, **62**: 40-49.

Albarrán, P., J. Crespo, I. Ortuño, and J. Ruiz-Castillo (2011), "The Skewness of Science In 219 Sub-fields and a Number of Aggregates", *Scientometrics*, **88**: 385-397.

Albarrán, P., Li, Y., and Ruiz-Castillo, J. (2012), "The Role of Extreme Observations in Citation Analysis", mimeo.

Bornmann, L, and H-D. Daniel (2008), "What Do Citation Counts Measure", *Journal of Documentation*, **64**: 45-80.

Bourguignon, F. (1979), "Decomposable income inequality measures", *Econometrica*, **47**: 901-920.

Braun, T., W. Glänzel, and A. Schubert (1985), *Scientometrics Indicators. A 32 Country Comparison of Publication Productivity and Citation Impact*, World Scientific Publishing Co. Pte. Ltd., Singapore, Philadelphia.

Cowell, F. A. and K. Kuga (1981), "Inequality Measurement: An Axiomatic Approach," *European Economic Review,* **15**: 287-305.

Herranz, N. & Ruiz-Castillo, J. (2012), "Multiplicative and Fractional Strategies When Journals Are Assigned to Several Sub-fields", in press *Journal of the American Society for Information Science and Technology* (DOI:10.1002/asi.22629).

Garfield, E. (1979), *Citation Indexing: Its Theory and Applications in Science, Technology, and Humanities*, New York: Wiley.

Glänzel, W. (2011), "The Application of Characteristic Scores and Scales to the Evaluation and Ranking of Scientific Journals", *Journal of Information Science*, **37**: 40-48.

Leydesdorff, L., and Opthof, T. (2010), "Normalization at the Field level: Fractional Counting of Citations", *Journal of Informetrics*, **4**: 644-646.

Moed H. F. (2010), "Measuring contextual citation impact of scientific journals", *Journal of Informetrics*, **4**: 265–77.

Moed, H. F., W. J. Burger, J. G. Frankfort, and A. F. J. van Raan (1985), "The Use of Bibliometric Data for the Measurement of University Research Performance", *Research Policy*, **14**: 131-149.

Moed, H. F., and A. F. J. van Raan (1988), "Indicators of Research Performance", in A. F. J. van Raan (ed.), *Handbook of Quantitative Studies of Science and Technology*, North Holland: 177-192.

Moed, H. F., R. E. De Bruin, and Th. N. van Leeuwen (1995), "New Bibliometrics Tools for the Assessment of national Research Performance: Database Description, Overview of Indicators, and First Applications", *Scientometrics*, **33**: 381-422.

Murugesan, P., and Moravcsik, M.J. (1978), "Variation of the Nature of Citation Measures with Journal and Scientific Specialties", *Journal of the American Society for Information Science and Technology*, **105**: 17268-17272.

Pinski, G. and Narin, F. (1976), "Citation Influencefor Journal Aggregates of Scientific Publications: Theory, with Applications to the Literature of Physics", *Information Processing and Management*, **12**: 297-312.

Radicchi, F., Fortunato, S., and Castellano, C. (2008), "Universality of Citation Distributions: Toward An Objective Measure of Scientific Impact", *PNAS*, **105**: 17268-17272.

Radicchi, F., and Castellano, C. (2012a), "A Reverse Engineering Approach to the Suppression of Citation Biases Reveals Universal Properties of Citation Distributions", *Plos One*, **7**, e33833, 1-7.

Radicchi, F., and Castellano, C. (2012b), "Testing the fairness of citation indicators for comparisons across scientific domains: The case of fractional citation counts", *Journal of Informetrics*, **6**: 121-130.

Ravallion, M., and Wagstaff, A. (2011), "On measuring scholarly influence by citations", *Scientometrics*, **88**: 321-337.

Roemer, J. (1998), *Equality of Opportunities*, Harvard University Press.

Ruiz-Castillo, J. (2003), "The Measurement of Inequality of Opportunities", in J. Bishop y Y. Amiel (eds.), *Research in Economic Inequality,* **9**: 1-34.

Shorrocks, A. F. (1980), "The Class of Additively Decomposable Inequality Measures," *Econometrica,* **48**: 613-625.

Shorrocks, A. F. (1984), "Inequality Decomposition by Population Subgroups," *Econometrica,* **52**: 1369-1388.

Schubert, A., and T. Braun (1986), "Relative Indicators and Relational Charts for Comparative Assessment of Publication Output and Citation Impact", *Scientometrics*, **9**: 281-291.

Schubert, A., and T. Braun (1996), "Cross-field Normalization of Scientometric Indicators", *Scientometrics*, **36**: 311-324.

Schubert, A., W. Glänzel, and T. Braun (1983), "Relative Citation Rate: A New Indicator for Measuring the Impact of Publications", in D. Tomov and L. Dimitrova (eds.), *Proceedings of the First National Conference with International Participation in Scientometrics and Linguistics of Scientific Text*, Varna.

Schubert, A., W. Glänzel and T. Braun (1987), "A New Methodology for Ranking Scientific Institutions", *Scientometrics*, **12**: 267-292.

Schubert, A., W. Glänzel, and T. Braun (1988), "Against Absolute Methods: Relative Scientometric Indicators and Relational Charts as Evaluation Tools", in A. F. J. van Raan (ed.), *Handbook of Quantitative Studies of Science and Technology*: 137-176.

Small, and Sweeney (1985), "Clustering the Science Citation Index Using Co-citations. I. A Comparison of Methods", *Scientometrics*, **7**: 393-404.

Vinkler, P. (1986), "Evaluation of Some Methods For the Relative Assessment of Scientific Publications", *Scientometrics*, **10**: 157-177.

Vinkler, P. (2003), "Relations of Relative Scientometric Indicators", *Scientometrics*, **58**: 687-694.

Waltman, L, and N. J. van Eck (2012), "Universality of citation distributions revisited", *Journal of the American Society for Information Science and Technology*, **63**:72-77.

Zitt M, and Bassecoulard, E. (1994), "Development of a Method for Detection and Trend Analysis of Research Fronts Built by Lexical or Cocitation Analysis", *Scientometrics*, **30**: 333-351.

Zitt M, and Small H. (2008), "Modifying the journal impact factor by fractional citation weighting: The audience factor", *Journal of the American Society for Information Science and Technology*, **59**: 1856-1860.

STATISTICAL APPENDIX

Table A. Number of Articles and Mean Citation Rates by Field

	Number of Articles	%	Mean Citation	Standard Deviation
A. LIFE SCIENCES	**1,806,398**	**40.4**		
1. Biology & Biochemistry	275,568	6.2	12.6	20.1
2. Clinical Medicine	947,261	21.2	9.7	21.6
3. Immunology	60,875	1.4	16.0	23.0
4. Microbiology	73,039	1.6	11.4	13.9
5. Molecular Biology & Genetics	122,233	2.7	20.4	32.7
6. Neuroscience & Behav. Science	140,686	3.2	13.7	18.2
7. Pharmacology & Toxicology	76,728	1.7	8.0	11.0
8. Psychiatry & Psychology	110,008	2.5	7.0	11.3
B. PHYSICAL SCIENCES	**1,282,919**	**28.7**		
9. Chemistry	550,147	12.3	7.6	14.2
10. Computer Science	98,727	2.2	3.0	13.8
11. Mathematics	117,496	2.6	2.5	5.2
12. Physics	456,144	10.2	6.9	14.9
13. Space Science	60,405	1.4	11.0	20.5
C. OTHER NATURAL SCIENCES	**1,150,428**	**25.7**		
14. Agricultural Sciences	82,837	1.9	4.9	7.2
15. Engineering	356,269	8.0	3.2	5.8
16. Environment & Ecology	109,826	2.5	7.1	10.3
17. Geoscience	120,059	2.7	6.7	10.0
18. Materials Science	199,364	4.5	4.5	8.9
19. Multidisciplinary	20,672	0.5	3.2	7.0
20. Plant & Animal Science	261,401	5.8	5.1	8.0
D. SOCIAL SCIENCES	**232,587**	**5.2**		
21. Economics & Business	63,380	1.4	4.0	7.1
22. Social Sciences, General	169,207	3.8	3.3	5.7
ALL FIELDS	**4,472,332**	**100**	**7.9**	**16.4**

Table B. Exchange Rates, Standard Deviations, and Coefficient of variation for the [356, 705] Interval

	Exchange Rates	Standard Deviation	Coefficient of Variation
	(1)	(2)	(3)
A. LIFE SCIENCES			
1. Biology & Biochemistry	18.1	1.0	0.053
2. Clinical Medicine	11.3	0.6	0.054
3. Immunology	23.8	1.9	0.078
4. Microbiology	18.1	1.4	0.079
5. Molecular Biology & Genetics	25.6	1.0	0.040
6. Neuroscience & Behav. Science	20.5	1.5	0.075
7. Pharmacology & Toxicology	11.6	0.9	0.078
8. Psychiatry & Psychology	8.8	0.8	0.091
B. PHYSICAL SCIENCES			
9. Chemistry	10.2	0.8	0.079
10. Computer Science	2.2	1.1	0.506
11. Mathematics	3.0	0.7	0.237
12. Physics	7.6	0.7	0.088
13. Space Science	13.7	1.0	0.072
C. OTHER NATURAL SCIENCES			
14. Agricultural Sciences	6.4	0.9	0.147
15. Engineering	3.7	0.6	0.167
16. Environment & Ecology	10.6	0.8	0.076
17. Geoscience	9.5	0.9	0.092
18. Materials Science	4.9	0.9	0.174
19. Multidisciplinary	2.4	1.1	0.472
20. Plant & Animal Science	7.0	0.6	0.092
D. SOCIAL SCIENCES			
21. Economics & Business	4.4	0.7	0.169
22. Social Sciences, General	3.9	0.6	0.165

Table 1. Citation Inequality Decomposition for Different Inequality Indices and Different Quantile Choices

Inequality	Quantile	Within-group	Skew. of Sc.	*ICP*	Total Citation	Percentages In %:		
Indices	Choice, Π	Term, W	Term, S	Term	Ineq., $I_a(C)$	(1)/(4)	(2)/(4)	(3)/(4)
		(1)	(2)	(3)	(4)	(5)	(6)	(7)
$a = 2$	10	1.0407	0.8547	0.2866	2.1820	47.7	39.2	13.13
	50	0.6817	1.1437	0.3570	2.1820	31.2	52.4	16.36
	100	0.5771	1.2275	0.3778	2.1820	26.4	56.2	17.31
	1,000	0.3072	1.4415	0.4334	2.1820	14.1	66.1	19.86
$a = 0, \varepsilon_1 = 0.1$	10	0.0702	0.8905	0.2093	1.1700	6.0	76.1	17.89
	50	0.0134	0.9237	0.2329	1.1700	1.2	79.0	19.90
	100	0.0063	0.9306	0.2331	1.1700	0.5	79.5	19.92
	1,000	0.0007	0.9273	0.2419	1.1700	0.1	79.3	20.68
$a = 0, \varepsilon_2 =$	10	0.1644	1.0168	0.4445	1.6258	10.1	62.5	27.34
	50	0.0316	1.1049	0.4893	1.6258	1.9	68.0	30.10
	100	0.0170	1.1093	0.4995	1.6258	1.0	68.2	30.72
	1,000	0.0017	1.1111	0.5129	1.6258	0.1	68.3	31.55
$a = 1, \varepsilon_1 = 0.1$	10	0.0914	0.6439	0.1120	0.8473	10.8	76.0	13.22
	50	0.0293	0.7024	0.1150	0.8473	3.5	83.0	13.58
	100	0.0188	0.7124	0.1154	0.8473	2.2	84.1	13.62
	1,000	0.0045	0.7265	0.1161	0.8473	0.5	85.8	13.70
ı $a = 1, \varepsilon_2 = 0.01$	10	0.0937	0.6613	0.1168	0.8721	10.7	75.9	13.40
	50	0.0296	0.7226	0.1202	0.8721	3.4	82.8	13.79
	100	0.0191	0.7328	0.1206	0.8721	2.2	84.0	13.83
	1,000	0.0046	0.7462	0.1211	0.8721	0.5	85.6	13.89
$a = 1$, ln(0)=0	10	0.0940	0.6636	0.1179	0.8755	10.7	75.8	13.46
	50	0.0300	0.7244	0.1211	0.8755	3.4	87.2	13.83
	100	0.0192	0.7348	0.1215	0.8755	2.2	83.9	13.88
	1,000	0.0046	0.7488	0.1221	0.8755	0.52	85.53	13.95

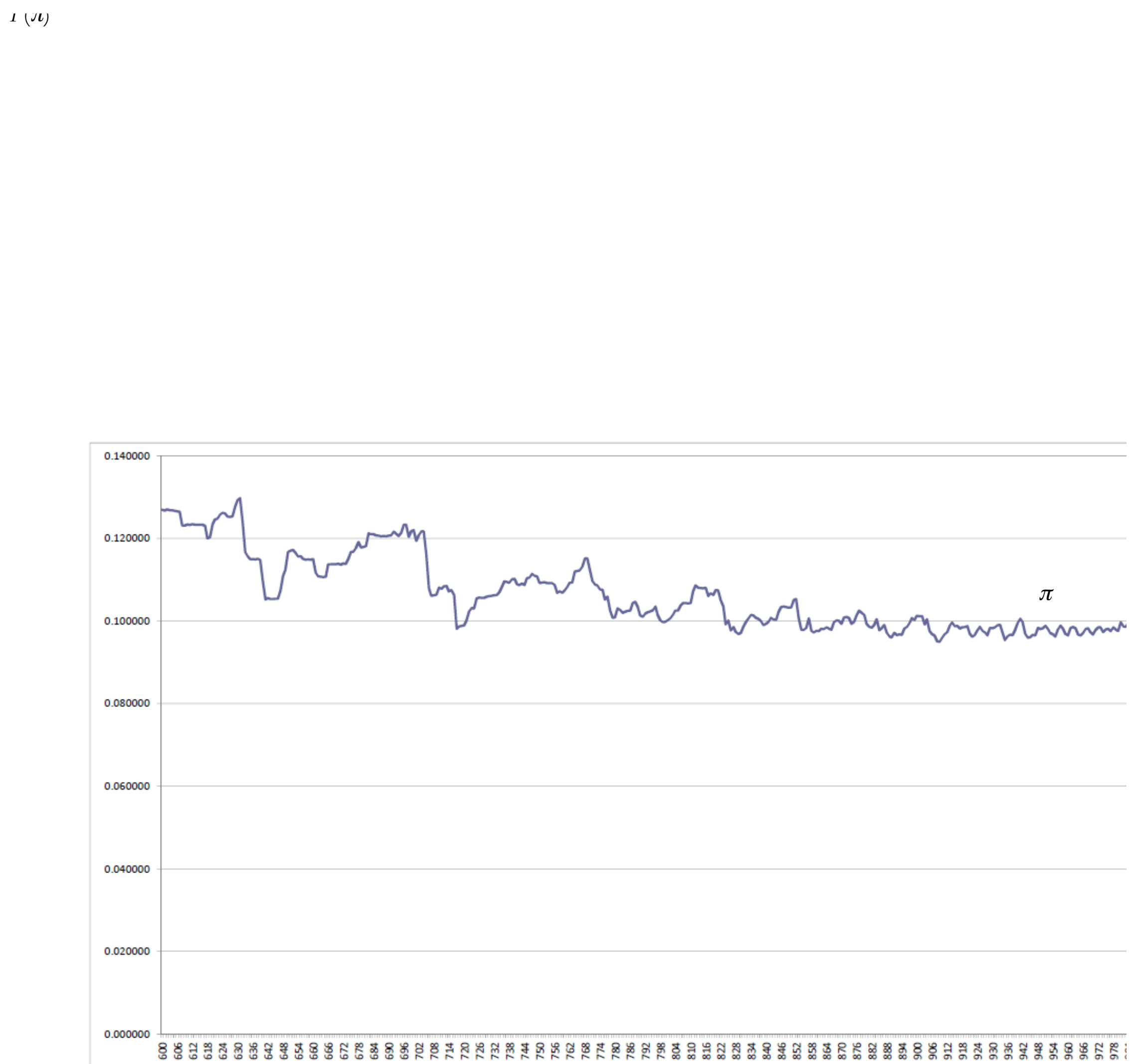

Figure 1. Citation Inequality Due to Differences in Citation Practices, $I(\pi)$ versus π. Raw Data

Table 2. Exchange Rates, Standard Deviations, and Coefficient of variation for the [706, 998] Interval, and Exchange Rates Based on Mean Citations

	Exchange Rates	Standard Deviation	Coefficient of Variation	% of Citations	ERs Based on Mean Citations	ERs Based on Mean Cits. In the [706, 998] Interval
	(1)	(2)	(3)	(4)	(5)	(6)
1. Biology & Biochemistry	15.8	0.9	0.054	68.0	16.0	15.3
2. Clinical Medicine	12.1	0.6	0.049	71.8	12.4	12.5
3. Immunology	19.5	0.9	0.048	66.3	20.4	19.0
4. Microbiology	14.4	1.3	0.092	65.8	14.6	13.5
5. Molecular Biology & Genetics	25.7	0.6	0.022	71.1	25.9	25.9
6. Neuroscience & Behav. Science	17.1	0.8	0.050	67.2	17.5	16.5
7. Pharmacology & Toxicology	10.1	0.6	0.056	68.4	10.2	9.8
8. Psychiatry & Psychology	9.1	0.2	0.025	72.4	9.0	9.1
9. Chemistry	9.9	0.4	0.037	70.9	9.7	9.7
10. Computer Science	3.7	0.5	0.124	76.3	3.8	4.0
11. Mathematics	3.3	0.2	0.059	75.4	3.1	3.3
12. Physics	8.8	0.5	0.061	74.2	8.7	9.1
13. Space Science	14.2	0.3	0.019	71.9	14.0	14.2
14. Agricultural Sciences	6.5	0.4	0.056	72.5	6.2	6.3
15. Engineering	4.4	0.2	0.054	75.9	4.1	4.4
16. Environment & Ecology	9.1	0.7	0.073	68.3	9.1	8.7
17. Geoscience	8.9	0.6	0.069	70.1	8.6	8.5
18. Materials Science	5.9	0.3	0.048	75.0	5.8	6.1
19. Multidisciplinary	4.3	0.7	0.158	81.6	4.1	4.7
20. Plant & Animal Science	6.7	0.3	0.045	71.3	6.5	6.5
21. Economics & Business	5.2	0.4	0.068	75.6	5.0	5.3
22. Social Sciences, General	4.5	0.2	0.045	75.1	4.2	4.5
Mean				**72.1**		

$_{v}^{\pi}I(\pi)$

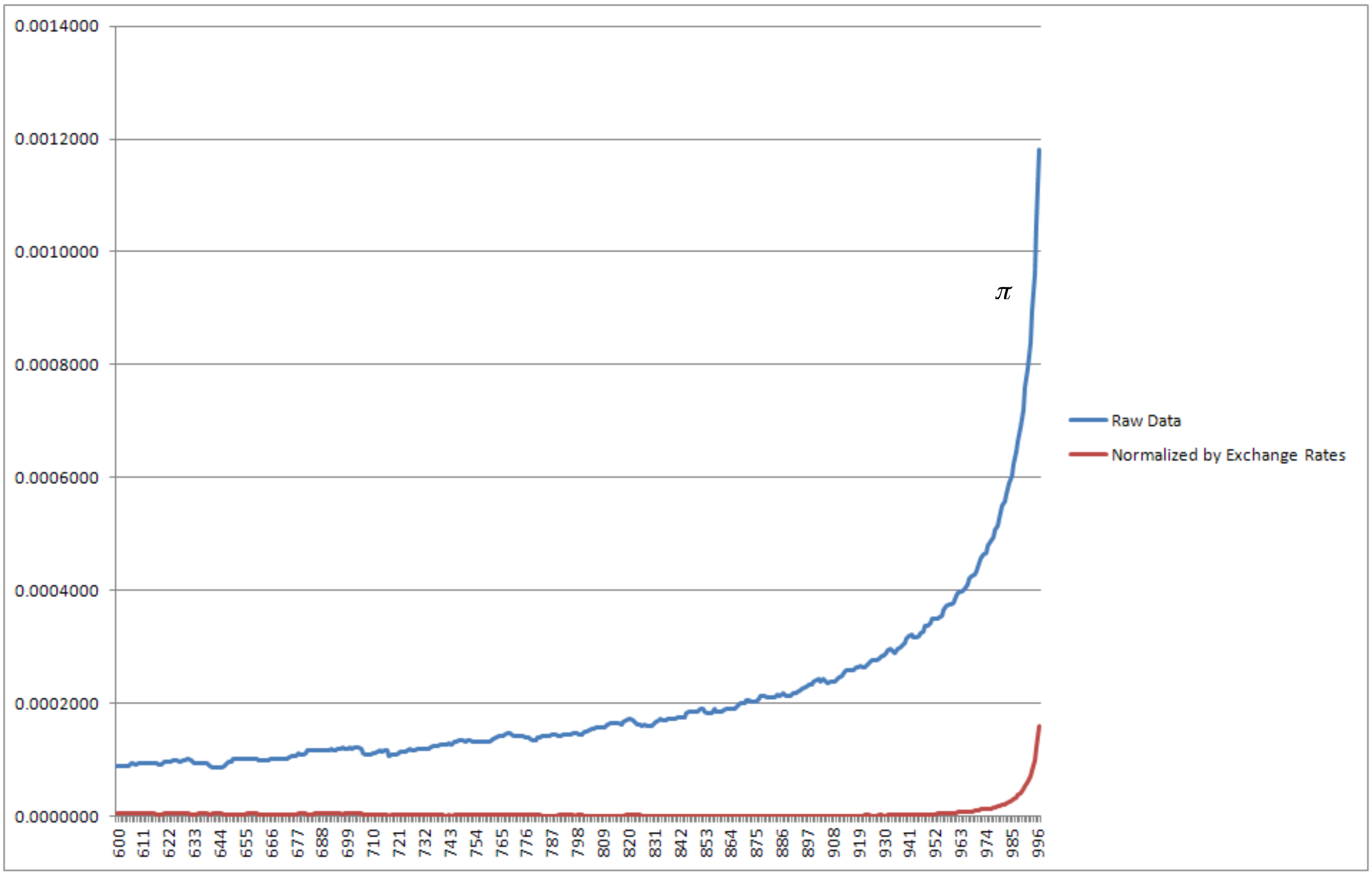

Figure 2. Weighted Citation Inequality Due to Differences in Citation Practices, $_{v}^{\pi}I(\pi)$ vs. π. Raw *vs.* Normalized Data

Table 3. Total Citation Inequality Decomposition Before and After Normalization: *IDCP* Interval Detail

	Within-group Term, W	Skew. of Sc. Term, S	*IDCP* Term	Total Citation Ineq., $I_1(C)$	Percentages In %: (1)/(4)	(2)/(4)	(3)/(4)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
A. RAW DATA							
All Quantiles	0.0046	0.7488	0.1221	0.8755	0.53	85.52	13.95
[1, 705]			0.0449				5.13
[706, 998]			0.0717				8.18
[999, 1000]			0.0056				0.64
B. EXCHANGE RATE NORMALIZATION							
All Quantiles	0.0051	0.7788	0.0167	0.8006	0.63	97.28	2.09
[1, 705]			0.0127				1.59
[706, 998]			0.0018				0.23
[999, 1000]			0.0022				0.27
C. NORMALIZATION WITH TWO EXCHANGE RATES							
All Quantiles	0.0050	0.7715	0.0147	0.7913	0.64	97.50	1.86
[1, 705]			0.0108				1.36
[706, 998]			0.0018				0.23
[999, 1000]			0.0021				0.27
D. MEAN NORMALIZATION							
All Quantiles	0.0050	0.7794	0.0164	0.8008	0.63	97.32	2.05
[1, 705]			0.0124				1.55
[706, 998]			0.0020				0.25
[999, 1000]			0.0020				0.25
E. GLÄNZEL NORMALIZATION							
All Quantiles	0.0048	0.7638	0.0241	0.7928	0.61	96.35	3.05
[1, 705]			0.0184				2.32
[706, 998]			0.0047				0.60
[999, 1000]			0.0010				0.13